DOCUMENT RESUME

ED 255 435 SO 016 333

AUTHOR Carlson, Ken
TITLE Teacher Evaluation in Social Studies.
PUB DATE Nov 84
NOTE 16p.; Paper presented at the Annual Meeting of the National Council for the Social Studies (64th, Washington, DC, November 15-19, 1984).
PUB TYPE Viewpoints (120) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Academic Achievement; Educational Policy; Elementary Secondary Education; *Evaluation Criteria; *Social Studies; *Teacher Evaluation

ABSTRACT

Student performance should be among the criteria used to evaluate teachers. Some advantages to comparing teachers on the basis of their students' learning are that such comparisons can provide (1) benchmarks by which to judge the validity of the criteria and measures of teaching effectiveness, (2) the evidence needed to match certain kinds of students with teachers whose style elicits maximum performance, and (3) the evidence needed to match team teachers with the areas in which they are most effective. In order for teachers to be evaluated fairly by such systems, the evaluation must be individualized to the students, the class, the particular subject taught, and the constraints of the environment in which teachers work. Teachers must be involved in designing the criteria rather than only making cosmetic changes in existing evaluation forms. Because of wide variations in goals and resources between state and local school districts, the evaluation should also be a localized procedure, occurring at the district level rather than the state or national level. (IS)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED255435

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

This document has been reproduced as received from the person or organization originating it.
Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
Kenneth Carlson

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

TEACHER EVALUATION IN SOCIAL STUDIES

Ken Carlson
Rutgers University

Paper Presented to the
College and University Faculty Assembly
of the
National Council for the Social Studies

Washington, D.C.
November 15, 1984

SO 016 333

INTRODUCTION

Good teachers are like pornography: I know 'em when I see 'em. Alas, other people refuse to agree with my perceptions because they are so damn perverse (not my perceptions - the other people). I have wearied of trying to bring enlightenment to these obtuse others, who seem to occupy an ever increasing proportion of the intellectual landscape. Given their dysgenic multiplication, I must assume that several of these philistines are in this very room right now. Thus, I shall not talk about the product of teacher evaluation but rather the process, about all one can do in a society which values pluralism more than Truth, and democracy more than wisdom.

My concern in this presentation will be with the essential ingredients of <u>any</u> procedure for evaluating social studies teachers. These ingredients are especially salient in "forced choice" evaluation, that is, evaluation which results in someone getting or losing a goody that is too scarce to be given to everyone, e.g., merit pay or master teacher status. The absence of the ingredients in most of the present evaluation systems is because these systems are not "forced choice"; everyone gets the goody, be it reappointment, tenure, or a salary increment. These "non-forced choice" systems, not so incidentally, are prevalent because they are popular with all the parties thereto. No one has to deny anything to anyone else, so no one is called upon to justify the denial and no one else feels denied. Everyone floats.

Of course, the impression this leaves with the public is that anyone can teach, and teach well, so hefty remuneration is not required to attract an adequate supply of laborers to this vineyard.

DECIDING WHO DECIDES

What, then, are the ingredients that should be common to all teacher evaluation systems? First, there is the matter of deciding who decides the criteria of effective teaching. Perhaps I've been in New Jersey too long (some would say that it's impossible to be in New Jersey too short), and maybe New Jersey is atypical, but the picture one gets over and over again from talking to social studies teachers in New Jersey is that they don't decide anything. They're not even sure who does decide. It's as though the decisions are being made by the No Man about whom James Herndon wrote in How to Survive in Your Native Land (1971). To be sure, collective bargaining rules in most states don't give teachers the right to negotiate much, but teachers aren't even asked for their opinions and the rules certainly do not prohibit consultation. A first order of business for social studies teachers, therefore, is to demand representation among the decisionmakers. Where state law, agency ruling, or court decree reserves the final decision to management, social studies teachers should at

least be among those who develop the proposals for management consideration.

This recommendatory role should entail more than reviewing the existing evaluation forms in the district and suggesting some changes in the wording. It must include the time and resources needed to review the literature on teacher evaluation as well as the evaluation systems being used in school districts which pride themselves on these. The Ford Foundation (1984) has noted that a major problem in city high schools, even in those that seem to be doing well, is lack of resources with which to evaluate teaching systematically. The literature review on teacher evaluation has already been facilitated by the Rand Corporation, which has made research on teacher evaluation one of its current priorities. (Darling-Hammond, Wise, & Pease, 1983) There are school districts which have achieved some acclaim for designing carefully their evaluation systems, and it is expected that the officials in each state education department would be able to name some of those districts in their state. Finally, there are the statewide teacher evaluation systems, such as that in Georgia.

In short, there is a wealth of ideas on how to evaluate teachers but no consensus. The lack of consensus makes a good case for district control and for the futility of trying to devise a system that will be adopted on a broader scale. States

may have the legal authority to impose a statewide system, but unless this system reflects local purposes and realities or is loose enough to accommodate such, it will be honored in the breach. Moreover, if states have the authority to mandate statewide teacher evaluation systems, they certainly don't have the personnel to administer these, so the most a state can do is spot monitor the local administration of the state system.

CRITERIA AND EVIDENCE

Having just asserted the inevitability of pluralism in teacher evaluation and the futility of trying to devise a system that will be acceptable universally, I will now embark on an exercise in futility. I won't attempt to specify the particular criteria of teaching effectiveness that should be used but only the kind of criteria. Traditionally, teaching effectiveness has been defined in terms of what teachers do. This has ranged all the way from the notorious example of keeping the window shades at even level to asking questions designed to elicit higher order thinking from pupils. And, of course, there have been the general homilies about being well-organized, fair and firm. When these exhortations are codified on an observation checklist, teachers usually manage to mimic the desired behavior well enough to pass muster or to incur no greater sanction than

an admonition to try harder. The next time the observer comes around, the behaviors that were found wanting the first time are demonstrated to the point of bludgeon. And even if they're not, the wholistic nature of the evaluation decision means that the teacher passes on balance, with perhaps another exhortation from the evaluator but this time one accompanied by a sigh of resignation. Moreover, as Joyce and Weil (1972) pointed out a dozen years ago, there is a "diverse range of alternative patterns upon which teachers may model their behavior . . . and there is no future in our persistent effort to describe 'best teaching practice'" (p. xi). (See also Hyman, 1984.) Even so, there are some teacher competencies which appear to be related logically to pupil performance, and a repertoire of these gives a teacher something to draw upon, just as it gives education professors something to teach. There must be thousands of methods books which deal with various teaching competencies, a fact which supports Joyce and Weil's pessimism. The point being made here is that competence should not be confused with effectiveness.

I would suggest that teacher evaluation be refocused from what teachers do to what it is they get their students to do. (Cf. Millman, 1981, 1982.) After all, schools exist for students. If the teacher looks good as a teacher should, but students display little if any learning, how can the teacher be

said to have served the real purposes of the school? One is struck by the fact that scholarly tomes on teacher competency, e.g., Felder (1978) and Funkhouser (1984), never treat teacher competency as being in any way related to, or measurable by, pupil performance. It is as though there are competent teachers and high achieving students, and any relationship between the two is purely coincidental. It is one thing not to confuse competence with effectiveness; it is quite another never to discuss the former in terms of the latter.

The kinds of student learning against which teachers should be judged will be determined by the local district, and this is where pluralism comes into play. Some districts might want students to excel on a test of American history names, places, and dates, God forbid. They may have a sense that this is what back-to-basics means. They may also think that it makes eminently good sense to test kids on the stuff that actually gets taught - you kow, like content validity. All that we sneered-at-savants can do in a case like this is continue working our wondrous powers of persuasion.

The reason that student-focused teacher evaluation is practically unheard of, despite the logic for doing it, is that it defies the use of a single evaluation form for all teachers of all subjects at all grade levels. Instead, it compels the use of course-specific evaluation. A teacher of U.S. History I and

a teacher of World History are not expected to effect the same student learning, so different measures of student learning are required for each. No more generic evaluation of nice teacher traits. The course-specific nature of student-focused teacher evaluation means, however, that it is virtually impossible to ignore the opinions of the course instructors in designing the evaluation instruments. If teachers are going to be judged in terms of what students are supposed to accomplish, they will demand a say in deciding what the expected student accomplishments are to be.

Whatever student learnings are decided on by the local district should be ranked in order of importance. It may not be possible to assign numerical weightings to these learnings, but there should be some idea of what teachers are most and least expected to accomplish with students. If some troglodyte on the school board insists on penmanship being taught across the curriculum, the most that can be done is to give it the least emphasis.

One kind of pupil outcome that should always be assessed is attitude toward the subject. If kids, under parent and teacher pressure, play the game and learn a lot but determine never to approach the subject again, future learning has been foreclosed, a situation that can be viewed with equanimity only if we assume that learning stops when schooling stops. Assessing

student attitude toward the subject is not the same as having students rate the instructor, something which is presently done mostly for college instruction and is fraught with problems even there. (Aleamoni, 1981)

The measures of student learning have to be behavioral. That is just another way of saying that student learning has to be measurable. The measures need not be paper-and-pencil tests but they should have the characteristics commonly associated with good measurement instruments, that is, validity and reliability. The opposite of measurement is mystification, the mumbo-jumbo of our profession with which we are less and less able to deflect a demanding public. There is a political reality to be reckoned with here: the public wants accountability in terms it can understand. (Sandefur, 1984) When teachers say that their greatest influence on students is like a long germinating plant that will take years to bloom, they are begging the question of whether that which they claim to have conceived will be a redwood or a weed. Why should the public give them the benefit of the doubt? Besides, if many years later the outgrowth happens to be a weed, how can the public then proceed to sanction a teacher who may have gone on to retirement - withhold his or her pension?

BEST COPY AVAILABLE

INSTRUCTIONAL VALIDITY

Let me here attempt to coin a phrase: "instructional validity." By this I mean that the kinds of student learning against which teachers are to be judged must be teachable. Not theoretically teachable but teachable in the circumstances in which the teachers find themselves. There is a lot of talk about teacher expectations for students, with the implication that the higher those expectations the better because students always live up (or down) to their teacher's expectations. Another school of thought holds that expectations should not be so high as to guarantee failure and frustration. Obviously, some delicate balancing is required to hit the optimum level of expectation. The same is true for the public's expectations of teachers. As Hazel Hertzberg (1982) has noted, a lot of the social studies reforms of the 1960s failed to take root in schools because there was too much being expected of teachers in the environments in which they work. Careful balancing must be done between the constraints under which teachers work and the kinds of effects we would like teachers to have.

If teachers are compared with each other on the basis of their students' demonstrated learnings, some idea can be had of the range of effects that is possible with a school's present range of teachers and students. For the comparison to

BEST COPY AVAILABLE

be fair, some preconditions must be established. Classes should be matched on the basis of pre-test results of student mastery of the learnings that the teachers are expected to impart. Absent this matching, there should at least be some analysis of covariance to adjust for initial class differences. There should also be some assurance that resources are available equally to all the teachers. Each teacher can decide whether and which resources to use, but for one teacher to have a monopoly over the single new class set of materials is hardly fair. Class sizes should be within a very narrow range. There are other conditions that it would be well to control for but the reality of schools won't permit this. For example, a class that meets first period or last period would seem to be at a natural disadvantage to a class that meets third period, but clearly the school day cannot begin and end with the third period. A class that meets close to the cafeteria during a lunch period would seem to be at a disadvantage to a class in a more removed location, but there is probably some law of physics which necessitates this arrangement.

Comparing teachers with each other on the basis of their students' demonstrated learning can be objected to on the prediction that it will lead to disruptive competition, especially if it is to result in the parceling out of limited goodies. It is even possible for this comparison to cause disruptive coop-

eration, whereby teachers agree to a maximum of effort and make life miserable for anyone who would be a ratebuster. These possibilities are empirical issues, as they say, and experimentation is the empirical test.

On the other hand, there are distinct advantages to comparing teachers on the basis of their students' learning. First, the comparison can provide benchmarks by which to judge the "instructional validity" of the criteria and measures of teaching effectiveness. Second, it can provide the evidence needed to match certain kinds of students with teachers whose style elicits maximum performance. (Cf. Santmire & Friesen, 1984) Third, it can even provide evidence with which to team teachers so that each is teaching the pupil learnings that he or she is most adept at accomplishing. (The Ford Foundation (1984) found that successful city high schools are ones in which the teaching staff acts as a team.) The last two of these possibilities - student-teacher matching and team teaching - would be made more feasible by the use of criterion-referenced testing, and NCSS Bulletin 64 offers some valuable suggestions in this regard. (Williams & Moore, 1980)

Finally, thought should be given to the desirability of post post-testing of students. Some learning washes out very quickly, and students who pass a test in June do not have the foggiest idea of what the answers are in August. If the school

BEST COPY AVAILABLE

district takes its criteria of student learning and teacher effectiveness seriously, it should want to know which teachers produce lasting results, even if the sticking period is only two months.

CONCLUSION

In conclusion, I would like to say that all the ideas for this presentation I got from my good friend and inspiration, Jack Nelson. The presentation got its genesis a month or so back when Jack took me aside and said, "Listen, kid, I wanna tell you a thing or two . . . "

REFERENCES

Aleamoni, Lawrence (1981). Student ratings of instruction. In Jason Millman (Ed.). Handbook of Teacher Evaluation. (Published in cooperation with the National Council on Measurement in Education.) Beverly Hills, CA: Sage Publications.

Darling-Hammond, Linda, Wise, Arthur, & Pease, Sara (1983). Teacher evaluation in the organizational context: A review of the literature. Review of Educational Research, 53(3), 285-328.

Felder, Dell (Ed.) (1978). Competency Based Teacher Education: Professionalizing Social Studies Teaching. (NCSS Bulletin 56). Washington, DC: National Council for the Social Studies.

Ford Foundation (1984). City High Schools: A Recognition of Progress. New York: Ford Foundation.

Funkhouser, Charles (Ed.) (1984) Teacher Education & Practice. (The Journal of the Texas Association of Colleges for Teacher Education). 1(1).

Herndon, James (1971). How to Survive in Your Native Land. New York: Bantam.

Hertzberg, Hazel (1982). Social studies reform: The lessons of history. In Irving Morrissett (Ed.). Social Studies in the 1980s. (A report of Project SPAN). Alexandria, VA: Association for Supervision and Curriculum Development.

Hyman, Ronald (1984). Testing for teacher competence: The logic, the law, and the implications. Journal of Teacher Education. 35(2), 14-18.

Joyce, Bruce, & Weil, M. (1972). Models of Teaching. Englewood Cliffs, NJ: Prentice-Hall.

Millman, Jason (1982). Research on Teacher Effectiveness: Implications for Teacher Assessment. Paper presented at the 12th Annual Large-Scale Assessment Conference. Boulder, CO.

Millman, Jason (1981). Student achievement as a measure of teacher competence. In Jason Millman (Ed.). Handbook of Teacher Evaluation. Beverly Hills, CA: Sage Publications.

BEST COPY AVAILABLE

Sandefur, J. T. (1984). Teacher competency testing: The public's mandate. In Charles Funkhouser (Ed.). Teacher Education & Practice. (The Journal of the Texas Association of Colleges for Teacher Education). 1(1), 11-16.

Santmire, Toni, & Friesen, Patricia (1984). A developmental analysis of research on effective teacher-student interactions: Implications for teacher preparation. In Egbert, Robert, & Kluender, Mary (Eds.). Using Research to Improve Teacher Education: The Nebraska Consortium. (Teacher Education Monograph No. 1). Washington, DC: ERIC Clearinghouse on Teacher Education.

Williams, Paul, & Moore, Jerry (Eds.) (1980). Criterion-Referenced Testing for the Social Studies. (NCSS Bulletin 64). Washington, DC: National Council for the Social Studies.

BEST COPY AVAILABLE